Chair for Algorithms
and Data Structures
Prof. Dr. Hannah Bast
Claudius Korzen

Information Retrieval
WS 2017/2018

http://ad-wiki.informatik.uni-freiburg.de/teaching

UNI FREIBURG

Exercise Sheet 8

Submit until Tuesday, December 19 at **12:00pm (noon)**

Exercise 1 (10 points)

Copy your code from Exercise Sheet 2 (or from the master solution for that sheet if you prefer) to a new folder *sheet-08*. Extend your code by a method *preprocessing_vsm* that builds the term-document matrix from the inverted index (using BM25 scores as entries).

Then add a method *process_query_vsm* such that the result list is obtained via a multiplication of the term-document matrix with the query vector (and not via merging of the inverted lists, like in Exercise Sheet 2).

Note 1: You can use the code from the lecture as an orientation. The only differences are that you should use BM25 scores, whereas in the lecture we used simple tf scores, and that you should use sparse matrices, because it would be much too slow (and too space consuming) otherwise.

Note 2: For debugging, it might be useful to keep your old *process_query* and *merge* methods around. However, please remove them in the final version of your submission: they will only make it harder for your tutor to understand your code and give you meaningful feedback.

Exercise 2 (10 points)

Re-run the evaluation of your system on the benchmark from Exercise Sheet 2, both with the original *process_query* (using the inverted index) and with the new *process_query_vsm* (using the term-document matrix). Make sure that the results are identical.

Re-run the evaluation with (at least three) different normalizations of the rows or columns of the term-document matrix, as explained in the lecture. Do this with BM25 scores and with ordinary tf scores. Which combination of score type and normalization gives the best result? Write your original results (from ES2) and your best new results (from this ES8) in a row in the table on the Wiki.

Add your code to a new sub-directory *sheet-08* of your folder in the course SVN, and commit it. Make sure that *compile*, *test*, and *checkstyle* run through without errors on Jenkins. Also commit the usual *experiences.txt* with your much appreciated brief and concise feedback.